Extension of Random Matrix Theory to the L-moments for Robust Portfolio Allocation

Ghislain Yanou

Documents de Travail du
Centre d'Economie de la Sorbonne

CES Working Papers

Extension of Random Matrix Theory to the L-moments for Robust Portfolio Allocation

Ghislain YANOU

2008.103

Maison des Sciences Économiques, 106-112 boulevard de L'Hôpital, 75647 Paris Cedex 13
http://ces.univ-paris1.fr/cesdp/CES-docs.htm
ISSN : 1955-611X

Extension of Random Matrix Theory to the L-moments for Robust Portfolio Allocation

Ghislain Yanou[†]

January 20, 2009

Abstract

In this paper, we propose a methodology for building an estimator of the covariance matrix. We use a robust measure of moments called L-moments (see Hosking, 1986), and their extension into a multivariate framework (see Serfling and Xiao, 2007). Random matrix theory (see Edelman, 1989) allows us to extract factors which contain real information. An empirical study in the American market shows that the Global Minimum L-variance Portfolio (GMLP) obtained from our estimator well performs the Global Minimum Variance Portfolio (GMVP) that acquired from the empirical estimator of the covariance matrix.

Keywords: Covariance Matrix, Lvariance-covariance, Lcorrelation, concomitance, Random matrix theory.

J.E.L. Classification: G.110, G.111.

---

[*]We are grateful to Thierry Chauveau, Emmanuel Jurczenko, Thierry Michel and Jean-Philippe Medecin for help and encouragement in preparing this work. The content of this article engages only its authors and does not necessarily reflect the opinions of their employers.

[†]University of Paris-1 (TEAM/CES/CNRS), 106 bd de l'hôpital F-75647 Paris cedex 13. E-mail: ghislain.yanou@gmail.com

Extension of Random Matrix Theory to the L-moments for Robust Portfolio Allocation

Résumé

Nous proposons dans cet article un estimateur robuste de la matrice de variance-covariance néecessaire à la construction des portefeuilles dont lunique objectif est la réduction de la volatilité. Dans cet optique, nous utilisons des moments alternatifs aux moments classiques appelés les L-moments (voir Hosking, 1986), et leurs extensions au cadre bi-varié (voir Serfling et Xiao, 2007). La théorie de la matrice aléatoire (voir Edelman, 1989) nous permet dextraire les facteurs qui contiennent réellement de linformation. Une étude empirique sur le marché américain montre que le portefeuille de L-variance minimale globale (obtenu à partir de la matrice de L-comoments dordre deux filtrés) a de meilleures performances que le portefeuille de variance minimale globale (obtenu à partir de lestimateur empirique de la matrice de variance-covariance).

Mots clés : Matrice de variance-covariance, L-moments, théorie de la matrice aléatoire.

Classification J.E.L. : G.110, G.111.

Extension of Random Matrix Theory to the L-moments for Robust Portfolio Allocation

1 Introduction

Markowitz (1952) showed that an investor who cares only about expected returns and volatility of static portfolio should hold a portfolio on the efficient frontier. To implement this portfolio in practice needs to estimate both expected returns and covariance matrix from the time series. Traditionally, the sample mean and the covariance matrix have been used for this purpose. But due to the estimations errors, the portfolio that relies on the sample estimate typically performs poorly out of sample.

It is well known that it is more difficult to estimate expected returns than covariance matrix (see Merton, 1980), and also that errors into the sample mean have a larger impact on portfolio weights than errors into the sample covariance matrix. For this reason, recent research has focused on the Global Minimum Variance Portfolio (GMVP), which relies solely on estimation of covariance, and thus, is less vulnerable to the estimation errors than the mean-variance portfolio. Indeed, the superiority of GMVP is highlighted by extensive empirical evidences which show that GMVP usually performs better out-of-sample than any other mean-variance portfolio, even when the Sharpe ratio or others performance measures that depend on both the portfolio mean and variance are used for evaluating performance[1].

However, as Pafka and Kondor (2004) state, the empirical estimator of the covariance matrix often suffers from the “curse of dimensions”. In practice, many times the length of the stock returns’ time series (T) used for the estimation is not big enough compared to the number of stocks (N) one wishes to consider. As a result, the obtained estimated covariance matrix is ill conditioned. Typically, an ill conditioned covariance matrix exhibits implausibly large off-diagonal elements. Michaud (1989) points out that inverting such a matrix amplifies the estimation errors tremendously. Furthermore, when N is bigger than T, the sample covariance matrix is even not invertible at all (see Ledoit and Wolf, 2003). Another limit of the empirical estimator of the covariance matrix is pointed out by DeMiguel and Nogales (2007), due to the fact that the empirical covariance matrix is

[1]Haugen, 1999 shows that GMVP from the S&P500 universe have a better sharpe ratio than the S&P500 index.

the maximum likelihood estimator of the covariance matrix. If the maximum likelihood estimators are very efficient for a normal distribution, they are very sensitive to deviation from the normal. Moreover, empirical evidences show the non-normality of returns on the market.

In the literature, several approaches have been proposed to deal with the problem of estimating the large number of elements in the covariance matrix. One approach is to impose some structure on the estimator of the covariance matrix by shrinking the empirical covariance matrix. Following Stein (1956), Ledoit and Wolf (2001) propose a weighted average estimator of the covariance matrix between the sample covariance and a target estimator well structured[2]. Fan, Fan and Jiang (2007) use a similar approach to give a stationary property to a time-domain[3] estimator of the covariance matrix.

A second approach consists to give some structural properties to the covariance matrix by imposing a portfolio norm constraint (see Frost and Savarino, 1988 and Chopra, 1993). DeMiguel *et al.* (2007) suggest to impose a norm constraint on the portfolio program and show some analytical relations between this constraint and the shortage threshold which can be supported by the investor.

Several empirical evidences call into question the one factor model and show that except the market factor, others risk factors exist and should be took into account (see Black, Jensen and Scholes, 1972), this is at the origin of the multi-factor models. Some statistical methods like the principal component analysis have been used by the literature to extract factors on the historical returns, but this approach do not allows distinguish between factors which contain real information and noise. Random matrix theory developed by physicians in order to understand the energy process for which sources are unknown (see Edelman, 1989), gives a solution for filtering noise. For an application of Random matrix theory to the portfolio asset allocation, see Laloux *et al.*, (1999) and Plerou *et al.* (2001).

Usually, the empirical variance is used to measure the portfolio volatility. But the classical variance tends to be very sensitive to extreme values notably when the size of the estimation window is not important in comparison with the number of assets in the universe. An alternative method to understand moments of a distribution is obtained by a linear combination of order statistics named L-moments. Introduced by Sillito (1951) and popularized by Hosking *et al.* (1985), L-moments can be interpreted, like classical moments, as simple descriptors of the shape of a general distribution and they offer a number of advantages over conventional moments. First, all of the population L-moments

[2]For instance, the one factor model of Sharpe (1963) allows to build a structured estimator of the covariance matrix.

[3]The time-domain estimators take into account all observed returns to build an estimator (the sample covariance matrix for instance), contrary to the state-domain estimators which consider only all historical data returns close of the actual return.

exist and determine uniquely a probability distribution, provided that the mean of the distribution exists (see Hosking, 1990). That is, a distribution may be specified by its L-moments, even if some of its conventional higher-order moments do not exist. Furthermore this specification is always unique. Second, their sample estimates are more robust to data outliers[4] and more efficient than classical moments (see Hosking, 1986). Moreover, although sample moment-based ratios can be arbitrarily large, sample standardized L-moments have algebraic bounds (see Hosking, 1989). Motivated by the sampling properties of L-statistics, Hosking and Wallis (1987) have advocated that the estimation method of L-moments must provide a better approximation of the unknown parent distribution than the traditional method of moments. Serfling and Xiao (2007) develop co-Lmoment in a multivariate framework and this makes interesting to use the Lvariance-covariance matrix in the portfolio allocation problem.

However, Jagannathan and Ma (2003) show that imposing a short sale constraint when minimizing the portfolio variance is equivalent to shrink the extreme elements of the covariance matrix. This simple remedy for dealing with estimations errors performs very well. In fact, Jagannathan and Ma (2003) find that the sample covariance matrix (with short sale constraints) performs almost as well as those constructed using robust estimators of the covariance matrix above. The goal of this paper is to propose an estimator of the covariance matrix which performs well than the empirical covariance matrix, even when a short sale constraint is imposed, by using the Random matrix theory to extract real information from the Lvariance-covariance matrix.

First, we propose a symmetric version of the Lvariance-covariance matrix for the Markowitz framework, then we show empirical evidences motivating the use of the Random matrix theory to extract factors which contain real information from the Lvariance-covariance matrix, finally an empirical study on the American market shows that the GMLP (Global Minimum Lvariance Portfolio) derived from our estimator performs well out-of-sample than the empirical covariance when a short sale constraint is imposed.

The remainder of this paper is organized as follow. Section two presents the L-moments and their multivariate extension. In section three, we introduce Random matrix theory, and explain the intuition behind their use in finance. Empirical evidences allows us to justify the use of Random matrix theory to extract factors which contain real information in the Lvariance-covariance matrix, this is the fourth section. Finally, in section five, we compare the GMLP derived from our estimator, with the GMVP derived from the empirical estimator.

[4]Since they are only linearly influenced by large deviations

2 Multivariate L-moments Definitions

2.1 L-moments Definitions and basic properties

The univariate L-moments can be defined as probability weighted moments, expectations of order statistics or as a covariance.

2.1.1 L-moments as Probability Weighted Moments

Grennwood *et al.* (1979) introduce probability weighted moments PWM defined by the following expression:

$$PWM_{p,r,s} = E\left[X^p \left\{F\left(X\right)\right\}^r \left\{1 - F\left(X\right)\right\}^s\right] \tag{1}$$

where X denotes a random variable and $F\left(.\right)$ the corresponding cumulative distribution function. When r is equal to one and s is null, we have a new expression of the probability weighted moments:

$$PWM_{p,1,0} = E\left[X^p \left\{F\left(X\right)\right\}\right] \tag{2}$$

which corresponds to the traditionnal moments of order p. L-moments are obtained by setting p equals one and s equals zero. We obtain the following expression:

$$\begin{aligned} \beta_r\left(X\right) &= E\left[X\left\{F\left(X\right)\right\}^r\right] \\ &= \int_0^1 x\left(u\right) u^r du \end{aligned} \tag{3}$$

where $x\left(u\right)$ denotes quantile of the cumulative distribution function. We define the L-moment of order k denotes λ_k,of the random variable X by the following expression :

$$\lambda_k\left(X\right) = \sum_{i=0}^{k-1} p^*_{k-1,i}\, \beta_i\left(X\right) \tag{4}$$

where:

$$P^*_{k,i} = \left(-1\right)^{k-i} \begin{pmatrix} k \\ i \end{pmatrix} \begin{pmatrix} k+i \\ i \end{pmatrix}$$

and $P^*_k\left(u\right)$ is the k^{th} shifted Legendre polynomial, related to the usual Legendre polynomials $P_k\left(u\right)$ by $P^*_k\left(u\right) = P_k\left(2u-1\right)$.

2.1.2 L-moments as Expectation of Order Statistics

Let $X_{1:N} \leq X_{2:N} \leq ... \leq X_{k:N}$ denote the order statistics of the random sample X of size N. L-moment of order k can be expressed as a combination of the expected order statistics:

$$\lambda_k(X) = k^{-1} \sum_{i=0}^{k-1} (-1)^i \begin{pmatrix} k-1 \\ i \end{pmatrix} E(X_{k-i:k}) \tag{5}$$

where $E(X_{k-i:k})$ denotes the expectation of order statistics:

$$E(X_{r:k}) = \frac{k!}{(r-1)!(k-r)!} \int_0^1 x(u) u^{r-1} (1-u)^{k-r} du \tag{6}$$

2.1.3 L-moments as a Covariance

Following the L-moment's representation as probability weighted moments, we re-express L-moments as covariance:

$$\lambda_k(X) = \sum_{i=0}^{k-1} p^*_{k-1,i} \beta_i(X)$$

with:

$$\beta_r(X) = E[X\{F(X)\}^r]$$

Since $p^*_0(.) \equiv 1$ and using the orthogonality property of functions p^*_k, we have a new expression of L-moments:

$$\lambda_k(X) = cov\left(X, p^*_{k-1}(F(X))\right) + 1_{\{k=1\}} E(X) \tag{7}$$

where $cov(.)$ denotes the covariance between the random variable X and the corresponding probability distribution $F(X)$. For $k = 2$ we find the following expression:

$$\lambda_2(X) = 2cov(X, F(X)) \tag{8}$$

which corresponds to the simple Gini mean difference (see Gini, 1912). The following picture shows the robust property of the second L-moment to the extreme returns in comparison with variance:

- Please, insert somewhere here Figure 1 -

2.2 The Lvariance-covariance Matrix

In a multivariate framework, the Gini mean difference corresponds to the following expression:

$$\lambda_2(X, Y) = 2cov(X, F(Y)) \tag{9}$$

where Y denotes a random variable of size N. Expression above corresponds to the second L-moment between the random variable X towards the random variable Y which is not the same than the second L-moment between the random variable Y towards the random variable X described by $\lambda_2(Y,X)$:

$$\lambda_2(Y,X) = 2cov(Y,F(X)) \tag{10}$$

That is, the Lvariance-covariance matrix $\hat{\mathbf{\Omega}}_{Lmom}$ between the multivariate random variables (X,Y) is obtained by the following expression:

$$\hat{\mathbf{\Omega}}_{Lmom} = \begin{pmatrix} \lambda_2(X) & \lambda_2(X,Y) \\ \lambda_2(Y,X) & \lambda_2(Y) \end{pmatrix} \tag{11}$$

and the derived Lcorrelation matrix $\hat{\mathbf{\Omega}}_{Lcorr}$ corresponds to the following expression:

$$\hat{\mathbf{\Omega}}_{Lcorr} = \begin{pmatrix} 1 & \tau_{X,Y} \\ \tau_{Y,X} & 1 \end{pmatrix} \tag{12}$$

where $\tau_{X,Y}$ and $\tau_{Y,X}$ are respectively the Lcorrelation coefficient between the random variable X towards the random variable Y, and the Lcorrelation coefficient between the random variable Y towards the random variable X with:

$$\begin{cases} \tau_{X,Y} = \dfrac{\lambda_2(X,Y)}{\lambda_2(X)} \\ \tau_{Y,X} = \dfrac{\lambda_2(Y,X)}{\lambda_2(Y)} \end{cases} \tag{13}$$

An important result about Lcorrelation is that like traditionnal version, its values lie between ± 1 (see Serfling and Xiao, 2007).

3 The Random Matrix Theory in Finance

The study of correlations between price changes of different stocks is of a scientific interest and of a practical relevance in quantifying the risk of a given stock portfolio. The problem is that although every pair of assets should interact either directly or indirectly, the precise nature of interaction is unknown. In some ways, the problem of interpreting the correlations between individual stock-price changes is reminiscent of the difficulties experienced by physicists in the fifties, in interpreting the spectra of complex nuclei. Large amounts of spectroscopic data on the energy levels were becoming available but were too complex to be explained by model calculations because the exact nature of the interactions were unknown. The Random matrix theory has been developed in this context, (see Wigner, 1956, Dyson, 1962, Dyson and Mehta, 1963, and Mehta 1991), to deal with the statistics of energy levels of complex quantum systems. With the minimal assumption of a random

Hamiltonian, given by a real symmetric matrix with independent random elements, a series of remarkable predictions were made and successfully tested on the spectra of complex nuclei. Deviations from the universal predictions of the Random matrix theory identify system-specific, non-random properties of the system under consideration, providing clues about the nature of the underlying interactions.

The Random matrix theory finds a justification in finance since the real process of the stock returns is unknown, that is the cross-correlation between stocks needs to be approached. Traditionally, the empirical estimators of the covariance matrix and the correlation matrix have been used in this context, but they contain much estimation errors (see Michaud, 1989), and we can expect that they are random for a large part. The idea behind the use of the Random matrix theory in finance comes from this observation, and the stake is to filter factors into the empirical correlation matrix, which have same properties than factors of a random matrix, under the null hypothesis[5]. That is, we can suppose that factors out of the null hypothesis contain real information. Laloux *et al.* (1999) show some empirical evidences justifying the use of the Random matrix theory in finance. Following the Edelman's thesis (1989), Plerou *et al.* (2001) perform a study of the Random matrix theory to understand cross-correlation of the high frequency financial returns. A recent work on the Random matrix theory applied in finance comes from Potters *et al.* (2005) and Conlon *et al.* (2008).

What is then the spectrum of a random correlation matrix? The answer is known due to the work of Marcenko and Pastur (1967). We consider an empirical correlation matrix $\mathbf{C}$ of N assets and T historical returns coming from an universe of returns characterized by X, we have:

$$\mathbf{C} = \frac{1}{N} X X^T \tag{14}$$

where X^T denotes the transpose of X. Let $\mathbf{R}$ be the random correlation matrix coming from a multivariate universe of Gaussian independent elements A of size $N \times T$, we have:

$$\mathbf{R} = \frac{1}{N} A A^T \tag{15}$$

By construction, $\mathbf{R}$ belongs the type of matrices often referred to Wishart matrices in multivariate statistics (see Wishart, 1928). Statistical properties of random matrices such $\mathbf{R}$ are known (see Dyson, 1971). Particularly, when $N \to \infty$ and $T \to \infty$, such that $q \equiv N/T$ is fixed, Segupta and Mitra (1999) show under the null hypothesis, the analytical distribution $P_{\mathbf{R}}(e)$ of its eingenvalues:

$$P_{\mathbf{R}}(e) = \frac{q}{2\pi} \frac{\sqrt{(e_+ - e)(e - e_-)}}{e} \tag{16}$$

[5]The null hypothesis states that the correlation matrix in the market is the identity matrix, what does not corresponds to the empirical evidence in the market.

where e denotes the eigenvalue bounded within e_- and e_+, with e_- and e_+ respectively the lowest and the largest eigenvalues of $\mathbf{R}$:

$$e_\pm = 1 + \frac{1}{q} \pm 2\sqrt{\frac{1}{q}} \tag{17}$$

We can expect that all eigenvalues of $\mathbf{C}$ coming from empirical returns X, within $[e_-, e_+]$ correspond to noise and have to be filtered, and eigenvalues which deviate from the theoretical spectrum contain real information and must be used to estimate correlation matrix. The following picture shows the density distribution of eigenvalues of stock returns in the S&P500 universe:

- Please, insert somewhere here Figure 2 -

Before apply the Random matrix theory to the Lcorrelation matrix, we need to show that their corresponding eingenvalues follow universal properties of a random Wishart matrix, this is the aim of the next section.

4 Investigation of Properties of Lcorrelation Matrix: Is It coherent with Random Matrix Theory?

The Wishart matrices, like traditionnal correlation matrix, are symmetrics contrary to the Lvariance-covariance matrix. Futhermore, the asset allocation process of Markowitz (1952), uses a quadratic equation to build the optimal portfolio and the estimator of the covariance matrix need to be symmetric in this context. In the next sub-section, we propose a methodology to transform the Lvariance-covariance matrix in a symmetric matrix.

4.1 Symmetric Version of the Lvariance-covariance Matrix

The Lvariance-covariance matrix characterizes the concomitance effects between two random variables, and is not necessary symmetric. For instance, the following picture shows the recursive evolution of the Lcovariance coefficients from Alcatel towards Siemens and from Siemens towards Alcatel:

- Please, insert somewhere here Figure 3 -

We see that the Lcovariance coefficient from Alcatel towards Siemens is not the same than the Lcovariance coefficient from Siemens towards Alcatel. This asymmetrical property of the L-moments characterize the concomitance effects between Alcatel and Siemens.

We propose a transformation of the Lvariance-covariance matrix in a symmetric matrix by preserving the asymmetrical concomitance effects. We propose the following formula for the Lvariance-covariance[6] matrix $\hat{\mathbf{\Omega}}_{Lmom}$:

$$\hat{\mathbf{\Omega}}_{Lmom} = \begin{pmatrix} \lambda_2(X) & \alpha_1[\lambda_2(X,Y)] + \alpha_2[\lambda_2(Y,X)] \\ \alpha_1[\lambda_2(X,Y)] + \alpha_2[\lambda_2(Y,X)] & \lambda_2(Y) \end{pmatrix} \tag{18}$$

where $(\alpha_i)_{i=1,2}$ denote respectively the weighted concomitance effects from the random variable X towards the random variable Y and the weighted concomitance effects from the random variable Y towards the random variable X:

$$\begin{cases} \alpha_1 = \dfrac{\lambda_2(X,Y)}{\lambda_2(X,Y)+\lambda_2(Y,X)} \\ \alpha_2 = \dfrac{\lambda_2(Y,X)}{\lambda_2(X,Y)+\lambda_2(Y,X)} \end{cases} \tag{19}$$

The following picture shows the recursive evolution for the symmetric version of the Lvariance-covariance matrix between Alcatel and Siemens:

- Please, insert somewhere here Figure 4 -

4.2 Distribution of Eigenvalues for the Lcorrelation Matrix

The single factor model of Sharpe (1963) only takes into account the market factor for understanding the cross-correlation in the market. We propose first to show adequacy of the Lvariance-covariance matrix with the Random matrix theory. Let the following model:

$$x_{it} = \alpha_i + \beta_i x_{mt} + \varepsilon_{it} \tag{20}$$

where parameters x_{it}, α_i, β_i, x_{mt} and ε_{it} denote respectively returns of asset i observed at t, liquidity factor of asset i, systematic risk of asset i, the market returns observed at t, and finally the residuals. We simulate in this controled process a $T \times N$ matrix of returns $(x_{it})_{(i,t)\in[1,\ldots,N]\times[1,\ldots,T]}$ by replacing the market returns x_{mt} by the S&P500 index returns. The following picture shows distribution of eigenvalues of the corresponding Lcorrelation matrix:

- Please, insert somewhere here Figure 5 -

It appears that all factors are positives and none is null, which supposed that the corresponding Lcorrelation matrix has an inverse. We also observe one factor which deviated

[6]The corresponding Lcorrelation matrix is not symmetric following our formula, however the derived Lcorrelation matrix is a regular matrix.

from the others. By construction, these others factors correspond to the noise because we run a single factor model. The theoretical upper bound e_+ of the random Wishart matrices equals 1.94 and there is only one eigenvalue higher than e_+ on the figure. The second largest eigenvalue (equals 1.39) is lower than e_+ and may be considered like noise. The theoretical lower bound e_- of the random Wishart matrices equals 0.37. It appears some eigenvalues lower than e_-, we will explain this observation later.

We now consider the real asset returns in the S&P500 universe, and represent distribution of eigenvalues of the corresponding Lcorrelation matrix, and the theoretical spectrum of the random Wishart matrices:

- Please, insert somewhere here Figure 6 -

It seems that distribution of eigenvalues of the Lcorrelation matrix have good agreement with the theoretical spectrum of the random Wishart matrices. The number of stocks considered in our database equals 207, that is e_{207} denotes the largest eigenvalue and e_1 the smallest eigenvalue. There are seven eigenvalues higher than e_+ which are e_{207}, e_{206}, e_{205}, e_{204}, e_{203}, e_{202}, and e_{201}. Eigenvalues within the theoretical distribution go from e_{200} to e_{66} and there are 65 (from e_1 to e_{65}) eigenvalues smallest than e_-. Plerou *et al.* (2001) show that eigenvectors corresponding to eingenvalues smaller than the theoretical lower bound e_-, contain as significant participants, pair of stocks which have the largest value of correlation coefficient in the data sample. In order to conclude that eigenvalues higher than e_+ (eigenvalues lower than e_+) contain real information (can be considered as noise), we need to find good agreement (none agreement) between universal properties of random Wishart matrices and eigenvalues of the Lcorrelation matrix from the S&P500 universe, lower than e_+ (higher than e_+). This is the aim of the next section.

4.3 Distribution of Eigenvector Components

Deviations of eigenvalues from the theoretical distribution $P_{\mathbf{R}}(e)$ suggest that they should also be displayed in the statistics of the corresponding eigenvector components (see Laloux *et al.* 1999). In this section, we analyze the distribution of the eigenvector components. The distribution $\left\{v_k^l; l = 1, ..., N\right\}$ of eigenvectors v_k for a random correlation matrix $\mathbf{R}$ have a Gaussian distribution with mean zero and unit variance:

$$\rho(v) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-v^2}{2}\right) \tag{21}$$

We propose in this sub-section to compare distribution of eigenvectors of the Lcorrelation matrix from the S&P500[7] universe, within and out of the theoretical distribution.

[7] We also called empirical Lcorrelation matrix.

For having good agreement between the Lcorrelation matrix and Random matrix theory, eigenvectors of the empirical Lcorrelation matrix within the theoretical distribution have to agree with a Gaussian distribution, and eigenvectors out of the theoretical distribution should not agree with a Gaussian distribution. We select two eigenvectors, the first v_{148} from eigenvalue $e_{-} \prec e_{148} \prec e_{+}$ and v_{207} from the largest eigenvalue $e_{207} \succ e_{+}$. We represent their distributions and compare them with the Gaussian distribution above. The following pictures show distribution of eigenvectors v_{148}, and v_{207}:

- Please, insert somewhere here Figure 7 -

We find good agreement between eigenvector v_{148} and the Gaussian distribution. Contrary to the distribution of eigenvector v_{207} which have extreme values deviating of the Gaussian distribution. We find for the other eigenvectors v_i from eigenvalues $e_{-} \prec e_i \prec e_{+}$ within the theoretical distribution good agreement with the Gaussian distribution. We illustrate on the following picture the *kurtosis* coefficients for all eigenvectors distribution:

- Please, insert somewhere here Figure 8 -

Distribution of eingenvectors at the center of picture have *kurtosis* coefficients almost equal to three, contrary to eigenvectors on the left and right edges. This suppose that, within the theoretical distribution, eigenvectors of the empirical Lcorrelation matrix have good agreement with eigenvectors of random Wishart matrices.

4.4 Interpretation of the Largest Eigenvalue and the Corresponding Eigenvector

Since all components participate in the eigenvector v_{207} corresponding to the largest eigenvalue e_{207}, we can hope that e_{207} represents the market factor. We quantitatively investigate this notion by comparing the projection (scalar product) of the time series x_t from the S&P500 universe on v_{207}, the corresponding market portfolio which is the S&P500 index. We compute projection $x_{207}(t)$ of the time series $x_j(t)$ on the eigenvector v_{207}:

$$x_{207}(t) \equiv \sum_{j=1}^{N} v_{207}^{j} x_j(t) \tag{22}$$

By construction, $x_{207}(t)$ is the portfolio returns defined by the largest eigenvalue e_{207}. In order to show that e_{207} corresponds to the market, we compare $x_{207}(t)$ with the S&P500 index. The following picture shows returns from $x_{207}(t)$ and from the S&P500 index:

- Please, insert somewhere here Figure 9 -

We find remarkably similar behavior between portfolio returns obtained from the largest eigenvalue and the S&P500 index. The empirical correlation coefficient between the two portfolios equals 0.94. We also compare $x_{148}(t)$ with the S&P500 index and find an empirical correlation coefficient equals 0.039. The following picture shows the strong correlation between $x_{207}(t)$ and the S&P500 index comparing to the weak correlation between $x_{148}(t)$ and the S&P500 index:

- Please, insert somewhere here Figure 10 -

The good agreement between x_{207} and the S&P500 index shows that the largest eigenvalue corresponds to the market factor. We propose in the next sub-section to study the other deviating eigenvalues.

4.5 Interpretation of the Other Deviating Eigenvalues

In order to study the other largest eigenvalues we need to remove the effect of the most largest eigenvalue e_{207} and construct a new Lcorrelation matrix. Following the one factor model above, we replace the market return x_{mt} by x_{207} and regress the universe returns:

$$x_{it} = \alpha_i + \beta_i x_{207} + \varepsilon_{it} \tag{23}$$

Using an ordinary least square regression, we estimate parameters α_i, β_i and the residuals ε_{it}. We build a new Lcorrelation matrix using the residuals. This Lcorrelation matrix not contain influence of the largest eigenvalue e_{207}. The following pictures show distribution of eigenvalues before and after removed influence of the largest eigenvalue:

- Please, insert somewhere here Figure 11 -

After influence of the largest eigenvalue has been removed, it seems that some eigenvalues which was firstly in the bulk, deviate now from the theoretical distribution. This phenomenon is mainly due by the fact that the largest eigenvalue by infuencing all stocks, imposes high Lcorrelation coefficients by pair of stocks. The following picture shows the distribution of Lcorrelation coefficients before and after removed contribution of the largest eigenvalue:

- Please, insert somewhere here Figure 12 -

We introduce now a measure coming from the localization theory (see Gurh *et al.* 1998) named inverse participation ratio (IPR), to quantify the number of significant participants of an eigenvector. For an eigenvector v_k^l, the corresponding IPR is defined as:

$$I_k = \sum_{l=1}^{N} \left(v_k^l\right)^4 \tag{24}$$

The meaning of IPR can be illustrated by two limiting cases: (i) a vector with identical components $v_k^l \equiv 1/\sqrt{N}$ has $I_k = 1/N$, whereas (ii) a vector with one component $v_k^l \equiv 1$ and the remainder zero has $I_k = 1$. That is, the IPR quantifies the reciprocal of the number of eigenvector components that contribute significantly. In the case (i), all components are equally took into account, the corresponding IPR equals $1/N$ and the inverse IPR (number of significant participants) equal to N. We use an identical approach to compute the number of significant participants of our eigenvectors. The following picture shows the number of significant participants by eingenvectors:

- Please, insert somewhere here Figure 13 -

We show that the largest eigenvalue influences on a large part of stocks, with a significant participants almost equals $1/I_{207} = 183$ for an universe of 207 stocks. This is the higher number of significant participants obtained. We also see that the smallest eigenvalues (corresponding to the eigenvalues which deviate from the theoretical distribution on the left edge) have the lowest number of significant participants[8]. The number of significant participants of eigenvectors obtained from the other deviating eigenvalues allows explaining them. For that, we now propose to analyze group of stocks influenced by the other deviating eigenvalues with the following process:

- We compute the IPR by eigenvectors obtained from the other deviating eigenvalues, and thus the corresponding number of significant participants,
- we then, choose a percentage for the number of stocks to consider among the significant participants. We obtain n_k stocks where k corresponds to the eigenvalue,
- for every other deviating eigenvalues, we select the n_k largest significant participants in their eigenvector components,
- finally, we perform a sectorial analysis of each significant participant selected in the previous step.

For instance, the number of significant participants for eigenvectors v_{206} and v_{205} are respectively $1/I_{206} \equiv 45$ and $1/I_{205} \equiv 63$. If we choose a percentage of 20%, the number of stocks to consider for eigenvectors v_{206} and v_{205} are respectively $n_{206} \equiv 9$ and $n_{205} \equiv 13$. That is, to interpret eigenvalue e_{206} (e_{205}), we only consider the nine (thirteen) largest components of eigenvalue v_{206} (v_{205}).

[8]This result differs of the observations of Plerou et *al.* (2001) which find large values of the inverse participation ratio at both edges of the theoretical distribution, suggesting a "random band" matrix structure.

We obtained for every other deviating eigenvalues a group of n_k stocks. The following picture shows market sectors of these stocks:

- Please, insert somewhere here Figure 14 -

We find that these eigenvectors partition the set of all stocks into distinct sectorial groups. We find sectorial groups which contains stocks of firms in utilities (v_{206}), stocks of firms in energy (v_{205}), a combination of healthcare and energy firms (v_{204}), information technology firms (v_{203}), stocks of financial firms (v_{202}) and finally stocks of consumer firms (v_{201})[9]. Plerou *et al.* (2001)[10] find that the second largest eigenvector[11] corresponds to large market capitalization firms. In the following table, we list by eigenvectors, the corresponding n_k firms with their corresponding sectors:

- Please, insert somewhere here Table 1 -

Concerning the smallest eigenvalues out of the theoretical distribution on the left edge, there is no evidence about a sectorial repartition. It seems that they group pair of stocks with homogeneous concomitance effects. In addition, their corresponding numbers of significant participants are low in comparison with other eigenvectors.

Our empirical observations seem to confirm expectation according to which eigenvalues higher than the theoretical upper bound e_+ contain real information, and eigenvalues smaller than e_+ can be considered as noise and have to be filtered. Since largest eigenvalues higher than e_+ contain real information, they characterize the market factors and we wish they are stable in the time. We investigate this point in the next section.

4.6 Stability of Eigenvectors in Time

Since they are the market components, we expect that eigenvectors obtained from the largest eigenvalues higher than e_+ are stable in the time. Let D_{jk} a matrix of size $p \times N$ defined as:

$$D_{jk} = \left\{ v_j^k; j = 1, ...p; k = 1, ..., N \right\} \tag{25}$$

where p denotes the number of eigenvalues higher than e_+. We next compute a matrix of size $p \times p$ named "overlap matrix" whose general term O_{ij} is defined as:

$$O_{ij}\left(t, \tau\right) = \sum_{k=1}^{N} D_{ik}\left(t\right) D_{ik}\left(t + \tau\right) \tag{26}$$

[9]Which is a mix between consumer staple and consumer discretionary.

[10]They use a more large universe of stocks in intradaily and daily frequencies.

[11]Corresponding in our study to e_{206}.

where t denotes initial time and τ future time. The "overlap matrix"defines the scalar product between eigenvectors from an initial time t to a future time τ. If all the p eigenvectors are perfectly non-random and stable in time we must have:

$$O_{ij}(t,\tau) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \tag{27}$$

The following picture illustrates "overlap matrix"obtained one week to sixteen years of smoothing windows[12]:

-Please, insert somewhere here Figure 15 -

Factors are plotted on the diagonal of the picture. At the beginning (when τ equals one week and for three years) on the left edge of the picture, we find good agreement between the initial sample and the future sample. We also see that only five factors deviated from the theoretical distibution. From the fourth year after the initial sample, only the two largest eigenvalues remain stable, and one year later we observe a sixth eigenvalue, and the second largest eigenvalue also becomes unstable. From the eleventh year after the initial sample, appears a seventh[13] eigenvalue and after fourteen year, the second largest eigenvalue is completely unstable and only the first largest eigenvalue which characterizes the market factor remains stable along the time. Out of the diagonal, the colour code seems shown that eigenvectors are almost perpendicular by pair.

Since the empirical Lcorrelation matrix have good agreement with Random matrix theory, the theoretical distribution of random Wishart matrices must be used to excract factors which contain real information in the Lcorrelation matrix, and then recover the Lvariance-covariance matrix. We now explain how to recover the Lvariance-covariance matrix from the Lcorrelation matrix.

5 The Filtered Lvariance-covariance Estimator of the Covariance Matrix

The idea consists in recovering from $\hat{\mathbf{\Omega}}_{Lmom}$ a new Lvariance-covariance matrix $\hat{\mathbf{\Omega}}_{FLmom}$ having the same trace. The following algorithm describes our methodology:

[12]We set the initial sample window from 05/29/1981 to 05/17/1991. We compute the "overlap matrix"between the initial sample window and a smooth window obtaind respectively one week later (from 06/05/1981 to 05/24/1991), one year later, two years later, until sixteen years later.

[13]When we consider the whole sample data, we find seven eigenvalues which deviate from the theoretical distribution.

- From the $T \times N$ matrix of returns, we first compute the symmetric version $\hat{\mathbf{\Omega}}_{Lmom}$ of the Lvariance-covariance matrix,
- we then compute the corresponding Lcorrelation matrix $\hat{\mathbf{\Omega}}_{Lcorr}$,
- next we compute the eigenvalues of $\hat{\mathbf{\Omega}}_{Lcorr}$ and for each eigenvalue, the percentage in the trace of $\hat{\mathbf{\Omega}}_{Lcorr}$,
- we identify the eigenvalues lower than the theoretical upper bound[14] e_+[15] and we set their values to zero,
- we then compute new values for the eigenvalues higher than theoretical upper bound e_+ from their corresponding percentage by preserving trace of $\hat{\mathbf{\Omega}}_{Lcorr}$,
- using new values of the eigenvalues, matrix of eigenvectors and its opposite, we compute the filtered Lcorrelation matrix $\hat{\mathbf{\Omega}}_{FLcorr}$,
- from $\hat{\mathbf{\Omega}}_{FLcorr}$, we compute the corresponding filtered Lvariance-covariance matrix $\hat{\mathbf{\Omega}}_{FLmom}$[16],

Finally we may use $\hat{\mathbf{\Omega}}_{FLmom}$ to estimate the covariance matrix. This way of doing is better than the empirical estimation of the covariance matrix with many respects. First, the L-moments are more robust than the standard moments. Second only real information is taken into account because noise has been filtered.

In the following section, we compare performances of GMLP (obtained from our estimator $\hat{\mathbf{\Omega}}_{FLmom}$) and GMVP (obtained from the empirical covariance matrix $\hat{\mathbf{\Omega}}_{Emp}$), when a short sale constraint is imposed.

[14] The theoretical upper bound is obtained from N and T.

[15] We neglect lowest eigenvalues because they have influence on a small number of stocks and produce none empirical evidences.

[16] Since $\hat{\mathbf{\Omega}}_{Lcorr} = \begin{pmatrix} 1 & \tau_{X,Y} \\ \tau_{Y,X} & 1 \end{pmatrix}$ where $\tau_{X,Y}$ and $\tau_{Y,X}$ correspond respectively to the Lcorrelation coefficient between the random variable X towards the random variable Y and the Lcorrelation coefficient between the random variable Y towards the random variable X from the symmetric version of the Lvariance-covariance matrix, with: $\begin{cases} \tau_{X,Y} = \dfrac{\lambda_2\,(X,Y)}{\lambda_2\,(X)} \\ \tau_{Y,X} = \dfrac{\lambda_2\,(X,Y)}{\lambda_2\,(Y)} \end{cases}$, we recover the Lvariance-covariance matrix from the following expression: $\hat{\mathbf{\Omega}}_{Lmom} = \begin{pmatrix} \lambda_2\,(X) & \tau_{X,Y} \times \lambda_2\,(Y) \\ \tau_{X,Y} \times \lambda_2\,(X) & \lambda_2\,(Y) \end{pmatrix}$.

6 An Application to the Portfolio Optimization

Jagannathan and Ma (2003) find that the sample covariance matrix (with short sale constraint) performs almost as well as those constructed using shrinkage estimators. The aim of our paper is to propose an estimator of the covariance matrix which performs well than the empirical covariance matrix, even when a short sale constraint is imposed. In this section, we perform an empirical study for comparing performances of the GMVP and the GMLP.

6.1 Portfolio Allocation Process

The optimization program when a short sale constraint has been set is given by:

$$\begin{cases} \underset{(\mathbf{w}_p)}{Min} \left(\mathbf{w}_p^{'}\mathbf{\Omega}\mathbf{w}_p\right) \\ s.t\ \mathbf{w}_p^{'}\mathbf{1} = 1 \\ \mathbf{w}_{p_I} \geq 0,\ i = 1, ..., N \end{cases} \tag{28}$$

Our database of origin is constituted of 207 assets of the S&P500 in a weekly frequency. In order to avoid in our optimization process, many weights close to zero, we propose to consider a new database. This new database is obtained from the significant participants for each eigenvalues which contains real information reported in Table 1 above. The number of assets in the new database is 65 from 05/29/1981 to 04/11/2008. The empirical protocol is the following:

- From the new databse, we consider data returns from 05/29/1981 to 05/23/1986, compute the optimal allocation and buy the corresponding portfolio,
- we then slide the estimation window for one week, that is we have a new estimation window from 06/05/1981 to 05/30/1986,
- next, we compute a new optimal allocation from the new estimation window,
- and we sell the old portfolio and buy a new portfolio corresponding to the new optimal allocation,
- we perform the algorithm until the 04/11/2008. Finally we obtain an out-of-sample portfolio from 06/05/1981 to 05/30/1986.

The following picture shows the net asset values of GMVP and GMLP in basis 100 along the estimation period:

- Please, insert somewhere here Figure 16 -

We also compute three statistic indicators for the out-of-sample portfolios which characterize the portfolio performances in terms of volatility, diversification and stability. For the portfolio performance's, we consider three statistics; the annualized standard deviation, the Sharpe ratioNo cash considered in our expression of the Sharpe ratio., and the tracking error. The formulas of all statistics above are available in appendix.

The following picture shows volatility of GMVP and GMLP returns' in an absolute framework and relative to the S&P500 index:

- Please, insert somewhere here Figure 17 -

The following table reports performance indicators of the GMVP, the GMLP and also for the corresponding market index (S&P500):

- Please, insert somewhere here Table 2 -

Concerning the portfolio diversification, we use the effective size which measures the effective number of assets which have been taken into account in the allocation process. If the optimal allocation is naive, the effective size is equal to N, and on the contrary, in the case where only one asset constitutes the optimal portfolio, the effective size is equal to one. The stability of the portfolio is measured by the turnover. The formulas of the effective size and turnover are available in appendix. The following pictures show effective size and turnover of the GMVP and the GMLP along the estimation period.

- Please, insert somewhere here Figure 18 -

6.2 Comments

It appears that the raw return of the GMLP is higher than the raw return of the GMVP, and is almost equal to the S&P500's index raw return. Concerning the volatility, difference between the annualized standard deviation of the GMVP and the GMLP is not relevant. It's seems that for an identical level of annualized standard deviation, our estimator allows to build a global minimum volatility portfolio[17] which has a relevant Sharpe ratio. Thus, the annualized mean return of the GMLP is higher of the one of the GMVP for about 150 basis points. Better, we see that the GMLP has a lower volatility relatively to the S&P500 index than the GMVP. This result supposes that the GMLP is a global minimum volatility portfolio which fits better with the market index. A similar result is found by Ledoit and Wolf (2004), which show that the relative volatility of a global minimum variance portfolio obtained from their shrinkage estimator of the covariance matrix is lower than the relative

[17]Do not confuse with the GMVP which is the Global Minimum Variance Portfolio.

volatility of the one from the empirical estimator of the covariance matrix, but they impose in their allocation program a less conservative short sale constraint.

An interesting result not reported here concerns the uncertainty relative to the out-of-sample strategy. We can measure this by computing the correlation coefficient between the out-of-sample portfolio and its corresponding in-sample portfolio[18]. We note a correlation coefficient of 0.92 between the in-sample and the out-of-sample GMVPs, and a correlation coefficient of 0.96 between the in-sample and the out-of-sample GMLPs. This result supposes that, our estimator have less uncertainty relative to the future than the empirical estimator of the covariance matrix.

Another interesting result of our estimator is the mean effective size obtained from the optimal weight along the estimation window. The GMLP have a mean effective size which is equal to 24% of the whole universe, that means that in average, 24% of the assets in the universe are effectively taken into account in the allocation process; it is equal to 17% for the GMVP. This result highlights the capacity of our estimator to diversify the optimal portfolio allocation. Thus, the GMLP is less sensitive to a specific stock than the GMVP, and the portfolio risk is diffused through a large number of assets.

The turnover measures the stability of the reallocation of the optimal portfolio between two estimation periods. The mean turnover of the GMVP is equal to 4.8% while the one of the GMLP is equal to 3.3%. This observation supposes that the pool of stocks take into account for the GMLP is more stable along time. A stable allocation process is important to reduce the transaction cost.

[18]The corresponding in-sample global minimum volatility portfolio is empirically the best portfolio which has the lowest volatility. It is used by practitioners for having an expected shape of their portfolio.

7 Conclusion

In this paper we propose a new estimator of the covariance matrix. We use an alternative method to understand moments of a distribution obtained from a linear combination of order statistics named L-moments. The Random matrix theory allows for extracting from the Lvariance-covariance matrix real information. Our aim is to build a Global Minimum Lvariance Portfolio (GMLP) which remains robust relatively to the Global Minimum Variance Portfolio (GMVP) obtained from the empirical estimator of the covariance matrix; even when a short sale constraint is imposed in the optimization process.

Furthermore, the asset allocation process of Markowitz (1952) uses a quadratic equation to build the optimal portfolio and the estimator of the covariance matrix need to be symmetric in this context. We propose a symmetric version of the Lvariance-covariance matrix.

In order to extract real information from the symmetric Lvariance-covariance matrix, we compare the theoretical distribution of eigenvalues of the random Wishart matrices with the distribution of the eigenvalues from the Lvariance-covariance matrix. This comparison requires in anticipation to find good agreements between universal properties of the random Wishart matrices and the Lcorrelation matrix. Some empirical evidences on the S&P500 universe confirm this point. We then extract eigenvalues from the Lcorrelation matrix which contain real information, and first we show that each one corresponds to a market sector of the S&P500 universe. Second, we show how to recover a filtered Lvariance-covariance matrix.

Finally, we compare the out-of-sample GMLP (obtained from the filtered Lvariance-covariance matrix) to the GMVP (obtained from the empirical estimator) when a short sale constraint has been set. Following our results, it seems that the GMLP outperforms the GMVP concerning the Sharpe ratio, the tracking error relatively to the S&P500 index, diversification and stability of the portfolio along time. Another interesting result is that, the uncertainty between the GMLP and its corresponding in-sample portfolio is lower than which obtained by the GMVP.

The methodology describes in this paper, may be also useful for practitioners which prefer selection than allocation, by considering only the first significant participants (stocks) which are described by each eigenvalues containing real information. A natural extension of this paper will be to perform a more advanced study on these stocks in order to highlight some style effects.

References

Bandi F.M. and P.C Phillips, (2003), “Fully Nonparametric Estimation of Scalar Diffusion Models”, *Econometrica 71*, 241-283.

Black F. , M.C. Jensen and M. Scholes, (1972), “The CAPM: Some Empirical Tests”, *in Jensen Studies in the Theory of Capital Markets, New York, Praeger Publishers*, 79-121.

Burda Z., A. Goerlich, A. Jarosz and J. Jurkiewicz, (2004), “Signal and Noise in Correlation Matrix”, *Physica A 343*, 295-317.

Chilson J., R. Ng, A. Wagner and R. Zamar, (2004), “Parallel Computation of High Dimensional Robust Correlation and Covariance Matrices”, *Proceedings of the ACM SIGKDD*, 533-538.

DeMiguel V., L. Garlappi, and R. Uppal,(2007), “Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy?”, *forthcoming in Review of Financial Studies*, 39 pages.

DeMiguel V., L. Garlappi, F.J. Nogales and R. Uppal, (2007), “A Generalized Approach to Portfolio Optimization: Improving Performance By Constraining Portfolio Norms”, *forthcoming in Management Science*, 36 pages.

Disatnik D. and S. Benninga, (2007), “Shrinking Covariance Matrix - The Simpler is Better”, *Journal of Portfolio Management 33*, 56-63.

Dyson F.J, (1962), “Statistical Theory of the Energy Levels of Complex Systems. I”, *Journal of Mathematics and Physics 3*, 140-156.

Dyson F.J. and M.L. Metha, (1963), “Statistical Theory of the Energy Levels of Complex Systems. V”, *Journal of Mathematics and Physics 4*, 713-719.

Edelman A. and R. Rao, (2005), “Random Matrix Theory”, *Acta Numerica 14*, 65 pages.

Edelman A., (1989), “Eigenvalues and Condition Numbers of Random Matrices”, *PhD Thesis, Department of Mathematics, Massachussetts Institute of Technology, Cambridge,* *MA*, 105 pages.

Fan, J., Y. Fan and J. Jiang, (2007), "Dynamic Integration of Time- and State-Domain Methods for Volatility Estimation", *Journal of American Statistical Association 102*, 618-631.

Fan, J., Y. Fan and J. Lv, (2007), "Aggregation of Nonparametric Estimators for Volatility Matrix", *Working paper, http://arxiv.org/abs/math.ST/0701107v1*, 46 pages.

Grubel R., (1988), "A Minimal Characterization of the Covariance Matrix", *Metrika 35*, 49-52.

Haugen R., (1999), "Inefficient Market", *Prentice Hall.*

Hosking J., J. Wallis and J. Wood, (1985), "Estimation of the Generalized Extreme-value Distribution by the Method of Probability-weighted Moments", *Technometrics 27*, 251-261.

Hosking J., (1986), "The Theory of Probability Weighted Moments", *Research Report RC12210, IBM Research Division*, 15 pages.

Hosking J. and J. Wallis, (1987), "Parameter and Quantile Estimation for the Generalized Pareto Distribution", *Technometrics 29*, 339-349.

Hosking J., (1989), "Some Theoretical Results Concerning L-Moments", *Research Report RC14492, IBM Research Division*, 13 pages.

Hosking J., (1990), "L-moments: Analysis and Estimation", *Journal of the Royal Statistical Society Series B 52*, 105-124.

Hosking J. and J. Wallis, (1997),"Regional Frequency Analysis", *Cambridge*, 224 pages.

Huang J., (1989), "Moment Problem of Order Statistics: A Review", *International Statistical Review 57*, 59-66.

Jagannathan R. and T. Ma, (2003), "Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps", *Journal of Finance 54*, 1651-1683.

Konheim T., (1971), “A Note on Order Statistics”, *American Mathematical Monthly 78*, 524.

Laloux L., P. Cizeau, J.-P. Bouchaud and M. Potters, (1999), “Noise Dressing of Financial Correlation Matrices”, *Physical Review Letter 83*, 1469.

Ledoit O. and M. Wolf, (2001), “Improved Estimation of the Covariance Matrix of Stock Returns With an Application to Portfolio Selection”, *Journal of Empirical Finance 10*, 603-621.

Ledoit O. and M. Wolf, (2003), “A Well-Conditioned Estimator for Large Dimensional Covariance Matrices”, *Journal of Multivariate Analysis 88*, 365-411.

Ledoit O. and M. Wolf, (2004), “Honey, I Shrunk the Sample Covariance Matrix”, *Journal of Portfolio Management 31*, 1-22.

Marcenko V. and L. Pastur, (1967), “Distribution of Eigenvalues for Some Sets of Random Matrices”, *Maths USSR Sborni 1*, 457-483.

Markowitz H., (1952), “Mean-Variance Analysis in Portfolio Choice and Capital Markets”, *Journal of Finance 7*, 77-91.

Markowitz H., (1959), “Portfolio Selection”, *Blackwell, Cambridge.*

Mehta M. L., (1991), “Random Matrices”, *Academic Press, Boston.*

Merton, R., (1980), “On Estimating the Expected Return on the Market: An Exploratory Investigation”, *Journal of Financial Economics 8(4)*, 323-361.

Michaud R., (1989), “The Markowitz Optimization Enigma: Is ’Optimized’ Optimal?”, *Financial Analysts Journal 45*, 31-42.

Michaud, R. (1998), “Efficient Asset Management: A Practical Guide to Stock Portfolio Optimization”, *Oxford University Press.*

Olkin I. and S. Yitzhaki, (1992), "Gini Regression Analysis", *International Statistical Review 60(2)*, 185-196.

Pafka S., M. Potters and I. Kondor, (2004), "Exponential Weighting and Random-Matrix-Theory-Based Filtering of Financial Covariance Matrices for Portfolio Optimization", *arXiv:cond-mat/0402573 v1 (Web-based journal).*

Plerou V., P. Gopikrishnan, B. Rosenow, L.A.N. Amaral, T. Guhr and H.E. Stanley, (2001), "A Random Matrix Approach to Cross-Correlations in Financial Data", *Physical Review E 65*, 19 pages.

Potters M., J.-P. Bouchaud and L. Laloux, (2005), "Financial Applications of Random Matrix Theory: Old Laces and New Pieces", *Acta Physica Polonica B 36(9)*, 2767-2784.

Serfling R. and P. Xiao, (2007), "A Contribution to Multivariate L-Moments: L-Comoment Matrices", *Journal of Multivariate Analysis, forthcoming*, 31 pages.

Shalit H. and S. Yitzhaki, (1984), "Mean-Gini, Portfolio Theory, and the Pricing of Risky", *Journal of Finance 5*, 21 pages.

Sharpe W., (1963), "A Simplified Model for Portfolio Analysis", *Journal of Management Science 9*, 277-293.

Sillitto G., (1951), "Interrelations between Certain Linear Systematic Statistics of Sample from any Continuous Population", *Biometrika 38*, 377-382.

Sillitto G., (1969), "Derivation of Approximants to the Inverse Distribution Function of a Continuous Univariate Population from the Order Statistics of a Sample", *Biometrika 56*, 641-650.

Stein C., (1956), "Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution", *Proceedings of the Third Berkeley Symposium on Mathematical and Statistical Probability, University of California, Berkeley 1*, 197-206.

Wigner E.P., (1956), "Results and theory of resonance absorption", *Oak Ridge National Laboratories Press*, page 59.

Yitzhaki S. and I. Olkin, (1991), "Concentration Indices and Concentration Curves", *in K. Mosler and M. Scarsini (eds.), Stochastic Orders and Decisions under Risk, Institute of Mathematical Statistics: Lecture-Notes Monograph Series 19*, 380-392.

8 Appendix

8.1 List of Tables

Table 1(a): Sectorials groups by deviating eigenvalues

e_{206}		e_{205}	
Company	Sectors	Company	Sectors
CONSTELL ENERGY	Utilities	SMITH INTL	Energy
INTEGRYS ENERGY GROUP	Utilities	ROWAN COMPANIES	Energy
XCEL ENERGY	Utilities	HALLIBURTON	Energy
DUKE ENERGY	Utilities	QUESTAR	Utilities
PUBL SVC ENTER	Utilities	APACHE	Energy
SOUTHERN	Utilities	NOBLE ENERGY	Energy
PROGRESS ENERGY	Utilities	CONOCOPHILLIPS	Energy
FPL GROUP	Utilities	MURPHY OIL	Energy
AM ELEC POWER	Utilities	SCHLUMBERGER	Energy
CONSOL EDISON	Utilities	HESS	Energy
xxxxxxx	xxxxxxx	OCCIDENTAL	Energy
xxxxxxx	xxxxxxx	EXXON MOBIL	Energy
xxxxxxx	xxxxxxx	CHEVRON	Energy

Table 1(a): **Source** : *Reuters, Sectorial groups of deviating eigenvectors* e_{206}*, and* e_{205}*, only the first* n_k *firms have been considered, from 207 assets of the S&P500 index, no completion need, from 05/22/1981 to 04/11/2008, weekly frequency, computation by the authors.*

Table 1(b): Sectorials groups by deviating eigenvalues

e_{204}		e_{203}	
Company	Sectors	Company	Sectors
PROCTER & GAMBLE	Healthcare	AMERICAN EXPRESS	Financials
APACHE	Energy	ADV MICRO DEV	Infotech.
BRISTOL MYERS	Healthcare	IBM	Infotech.
HALLIBURTON	Energy	CORNING	Infotech.
HJ HEINZ	Consumer	MOLEX	Infotech.
ELI LILLY	Healthcare	JPMORGAN CHASE AND CO	Financials
MURPHY OIL	Energy	HEWLETT PACKARD	Infotech.
HESS	Energy	TERADYNE	Infotech.
MERCK & CO	Healthcare	NATL SEMICONDUCT	Infotech.
EXXON MOBIL	Energy	MERRILL LYNCH	Financials
ABBOTT LABS	Healthcare	MOTOROLA	Infotech.
CONOCOPHILLIPS	Energy	ANALOG DEVICES	Infotech.
SCHLUMBERGER	Energy	TEXAS INSTRUMENT	Infotech.
PFIZER	Healthcare	xxxxxxx	xxxxxxx
CHEVRON	Energy	xxxxxxx	xxxxxxx
JOHNSON&JOHNSON	Healthcare	xxxxxxx	xxxxxxx

Table 1(b): **Source** : *Reuters, Sectorial groups of deviating eigenvectors* e_{204}*, and* e_{203}*, only the first* n_k *firms have been considered, from 207 assets of the S&P500 index, Infotech. denotes the Information Techology sector, no completion need, from 05/22/1981 to 04/11/2008, weekly frequency, computation by the authors.*

Table 1(c): Sectorials groups by deviating eigenvalues

e_{202}		e_{201}	
Company	Sectors	Company	Sectors
MARSH & MCLENNAN	Financials	GENERAL MILLS	Consumer
LENNAR CLASS A	Financials	DONNELLEY SONS	Industrials
AON	Financials	NEW YORK TIMES	Consumer
AMERICAN EXPRESS	Financials	WASHINGTON POST	Consumer
LINCOLN NATL	Financials	GANNETT	Consumer
TORCHMARK	Financials	CENTEX	Financials
CENTEX	Financials	MASCO	Industrials
JPMORGAN CHASE AND CO	Financials	CAMPBELL SOUP	Consumer
BANK OF NEW YORK	Financials	CONAGRA FOODS	Consumer
WELLS FARGO	Financials	WENDY'S INTL	Consumer
BOA	Financials	PULTE HOMES	Consumer
xxxxxxx	xxxxxxx	VARIAN MEDICAL	Healthcare
xxxxxxx	xxxxxxx	HERSHEY CO	Consumer

Table 1(c): **Source** : *Reuters, Sectorial groups of deviating eigenvectors* e_{202}, *and* e_{201}, *only the first* n_k *firms have been considered, from 207 assets of the S&P500 index, Consumer sector is a mix between Consumer Staple and Consumer Discretionary, no completion need, from 05/22/1981 to 04/11/2008, weekly frequency, computation by the authors.*

Table 2: Performance of the out-of-sample GMVP and GMLP

	GMVP	GMLP	S&P500 Index
Raw Return	434.00%	576.00%	597.00%
Annualized Mean Return	7.50%	9.00%	9.24%
Annualized Standard Deviation	9.80%	10.00%	13.29%
Sharpe Ratio	0.77	0.90	0.70
Tracking Error	9.00%	7.50%	xxxxxxx

Table 2: **Source** : *Reuters, Performances of the out-of-sample GMVP and GMLP, 260 periods for the sample window, 1142 periods of estimation, from 65 assets of the S&P500 index corresponding to the sectorial groups of deviating eigenvalues, no completion need, from 05/22/1981 to 04/11/2008, weekly frequency, computation by the authors.*

8.2 List of Figures

Figure 1: Recursive Variance and Lvariance

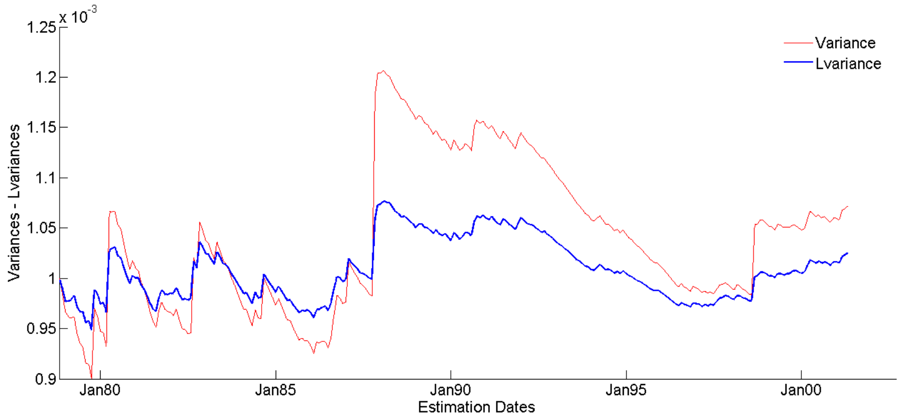

Figure 1: **Source** : *Reuters, S&P500 index, recursive variances and Lvariances, variance scales to the values of Lvariance, from 12/31/1974 to 04/30/2001, daily frequency, computation by the authors.*

Figure 2: Probability Density of Eigenvalues from the S&P500

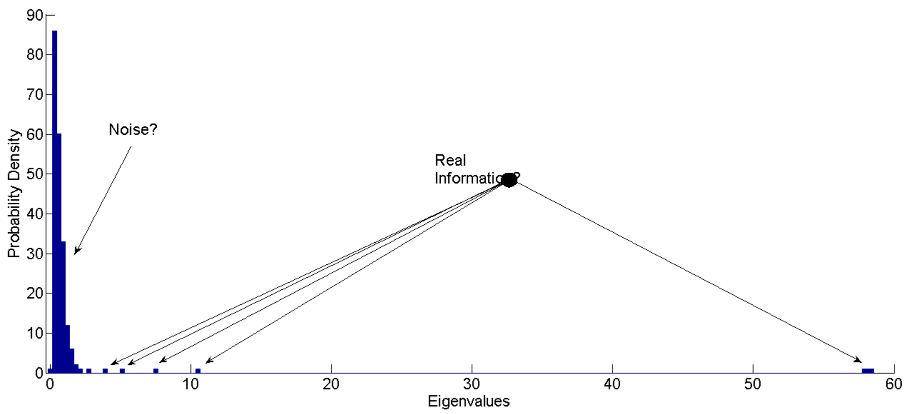

Figure 2: **Source** : *Reuters, distribution of eigenvalues from 207 assets, of the S&P500 index, no completion need, from 05/22/1981 au 04/11/2008, weekly frequency, computation by the authors.*

Figure 3: Recursive Lcovariance Coefficients

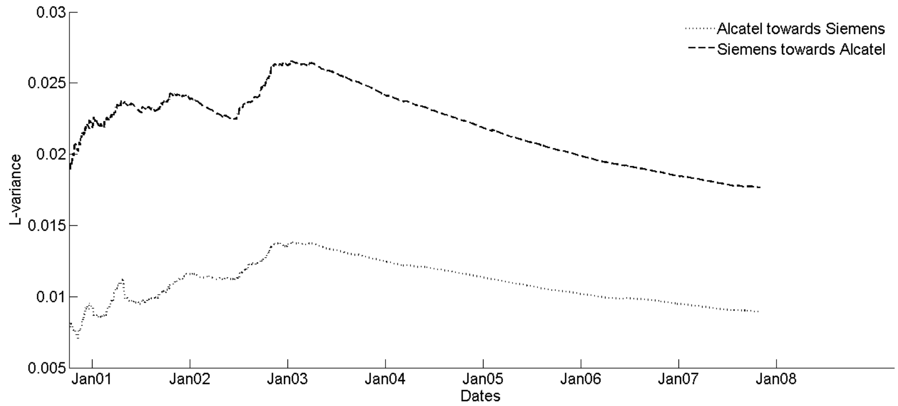

Figure 3: **Source** : *Reuters, Lvariance coefficients between two europeans stocks, from 11/04/2002 to 01/18/2008, daily frequency, no completion need, computation by authors.*

Figure 4: Recursive Lcovariance Coefficients: Symmetric Version

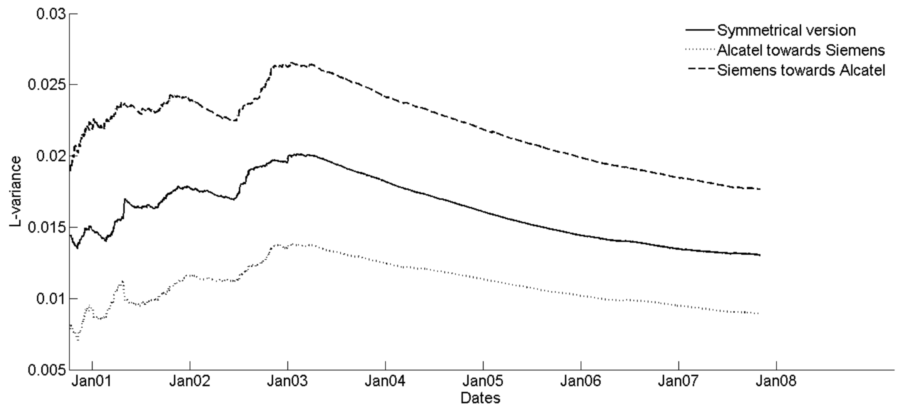

Figure 4: **Source** : *Reuters, Lvariance coefficients between two europeans stocks, from 11/04/2002 to 01/18/2008, daily frequency, no completion need, computation by authors.*

Figure 5: Probability Density of Eigenvalues from the Single Factor Model

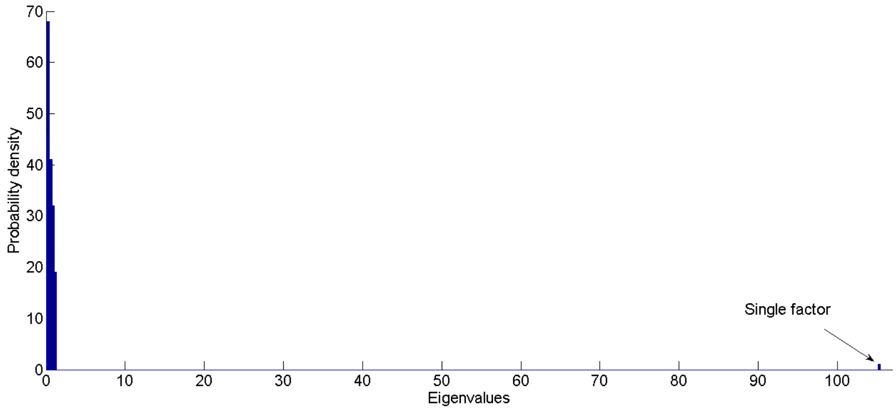

Figure 5: **Source** : *Reuters, distribution of eigenvalues for the single factor model, S&P500 like market, number of assets set to 207, number of historical returns set to 1402, from 05/29/1981 to 04/11/2008, weekly frequency, no completion need, computation by authors.*

Figure 6: Theoretical Probability Density of Eigenvalues

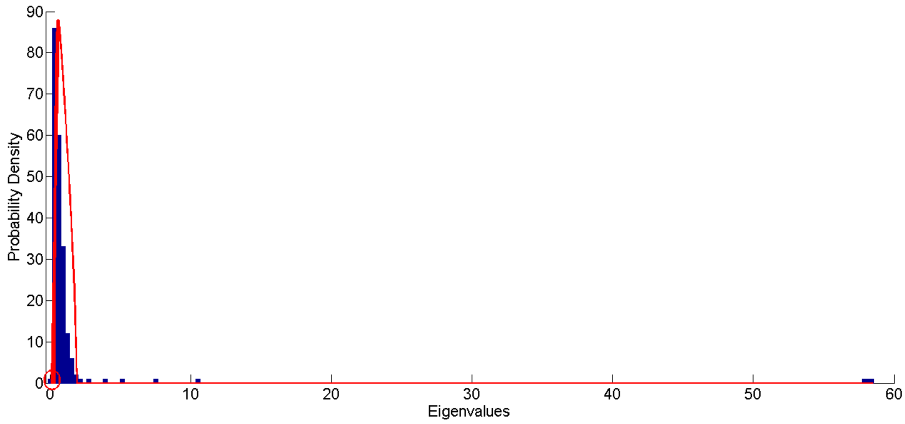

Figure 6: **Source** : *Reuters, distribution of eigenvalues from 207 assets of the S&P500 index, no completion need, from 05/22/1981 to 04/11/2008, weekly frequency, computation by the authors.*

Figure 7: Distribution of Eigenvector Components

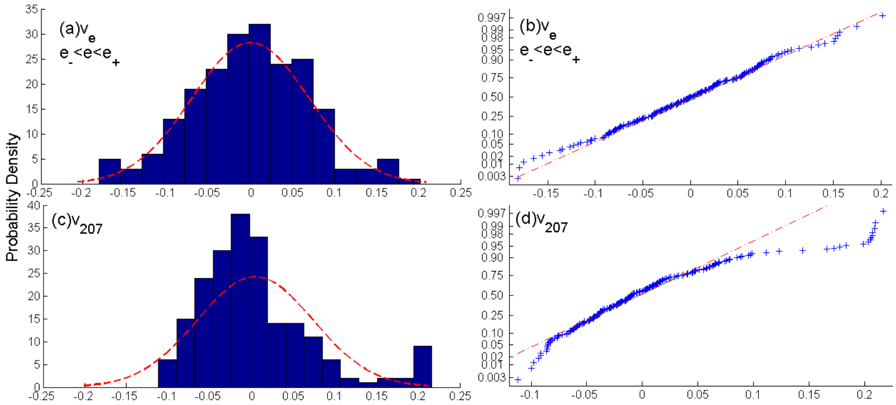

Figure 7: **Source** : *Reuters, comparison between distribution of eigenvectors v_{148} from eigenvalue e_{148} inside the theoretical distribution, and v_{207} from the largest eigenvalue e_{207}, with a Gaussian distribution in dashed, (a) and (b) represent distribution of v_{148}, (c) and (d) represent distribution of v_{207}, from 207 assets of the S&P500 index, no completion need, from 05/22/1981 to 04/11/2008, weekly frequency, computation by the authors.*

Figure 8: *Kurtosis* Coefficients from the Eigenvector Components

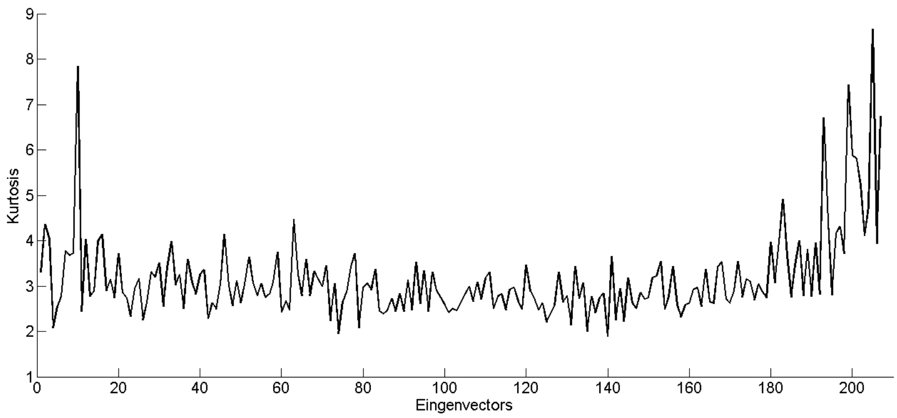

Figure 8: **Source** : *Reuters, kurtosis of the distribution of the whole eigenvectors, from 207 assets of the S&P500 index, no completion need, from 05/22/1981 to 04/11/2008, weekly frequency, computation by the authors.*

Figure 9: The Largest Eigenvalue Portfolio and the S&P500 index

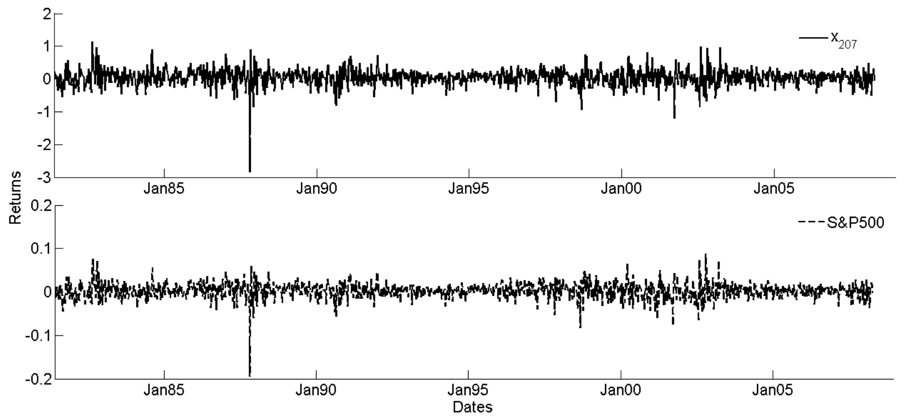

Figure 9: **Source** : *Reuters, comparison between the S&P500 index and the returns* x_{207} *coming from the largest eigenvalue* e_{207}*, from 207 assets of the S&P500 index, no completion need, from 05/22/1981 to 04/11/2008, weekly frequency, computation by the authors.*

Figure 10: Correlation Between Eigenvalue Portfolios and the S&P500 Index

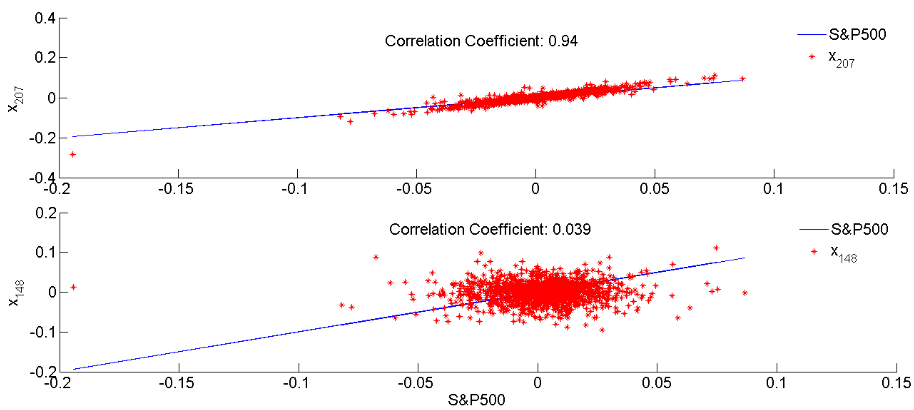

Figure 10: **Source** : *Reuters, correlation between the S&P500 index and the returns* x_{207} *coming from the largest eigenvalue* e_{207}*, and the returns* x_{148} *coming from an eigenvalue inside the theoretical distribution, from 207 assets of the S&P500 index, no completion need, from 05/22/1981 to 04/11/2008, weekly frequency, computation by the authors.*

Figure 11: Probability Density of Eigenvalues from the S&P500 Without Contribution of the Market Factor

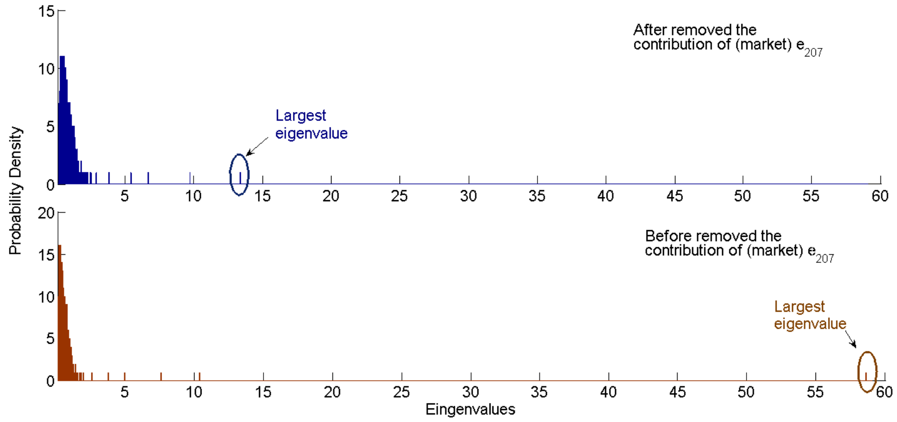

Figure 11: **Source** : *Reuters, comparison of the distribution of eigenvalues before and after removed influence of the largest eigenvalue* e_{207}*, from 207 assets of the S&P500 index, no completion need, from 05/22/1981 to 04/11/2008, weekly frequency, computation by the authors.*

Figure 12: Distribution of Lcorrelation Coefficients Without Contribution of the Market Factor

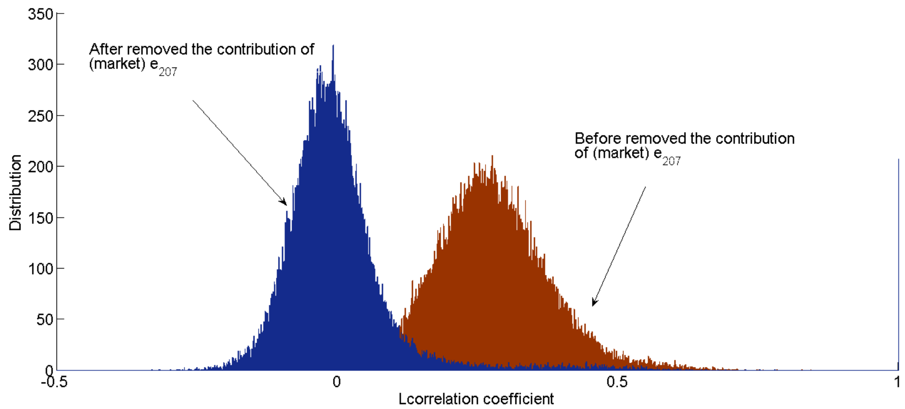

Figure 12: **Source** : *Reuters, Lcorrelation distribution of the universe before and after removed influence of the largest eigenvalue* e_{207}*, from 207 assets of the S&P500 index, no completion need, from 05/22/1981 to 04/11/2008, weekly frequency, computation by the authors.*

Figure 13: Number of Significant Participants by Eigenvectors

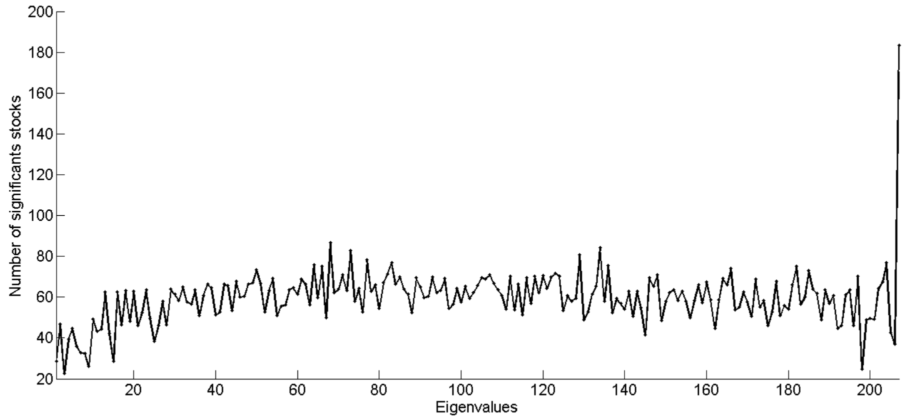

Figure 13: **Source** : *Reuters, number of significant participants by eigenvectors after removed influence of the largest eigenvalue* e_{207}*, from 207 assets of the S&P500 index, no completion need, from 05/22/1981 to 04/11/2008, weekly frequency, computation by the authors.*

Figure 14: Sectorial Repartition of Firms

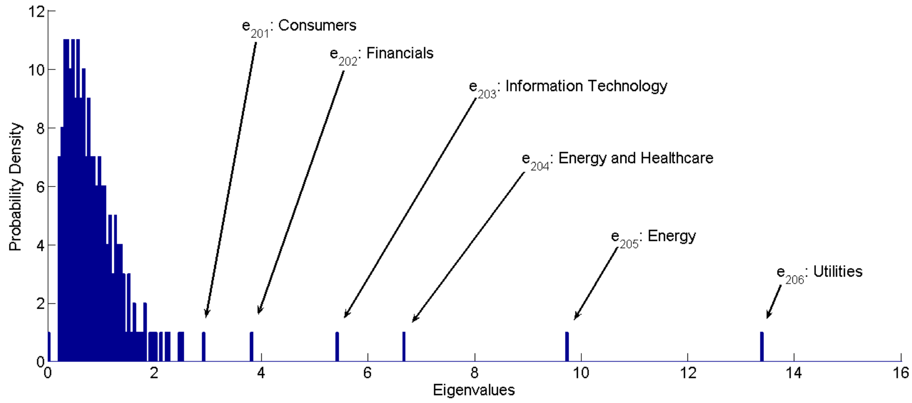

Figure 14: **Source** : *Reuters, sectorial groups of firms for eigenvalues* e_{206} *to* e_{201}*, from 207 assets of the S&P500 index, no completion need, from 05/22/1981 to 04/11/2008, weekly frequency, computation by the authors.*

Figure 15: Stability of Eigenvalues in the Time

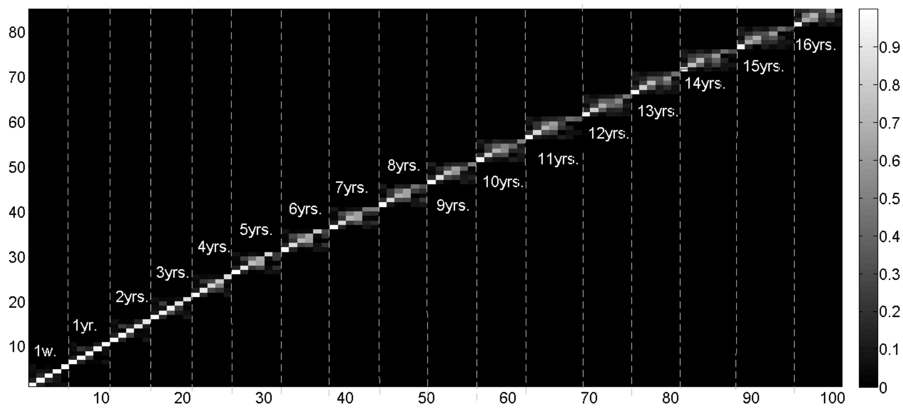

Figure 15: **Source** : *Reuters, stability of deviating eigenvalues through time, 520 periods for the initial sample, from 207 assets of the S&P500 index, no completion need, from 05/22/1981 to 04/11/2008, weekly frequency, computation by the authors.*

Figure 16: Returns of GMVP, GMLP and the S&P500 Index

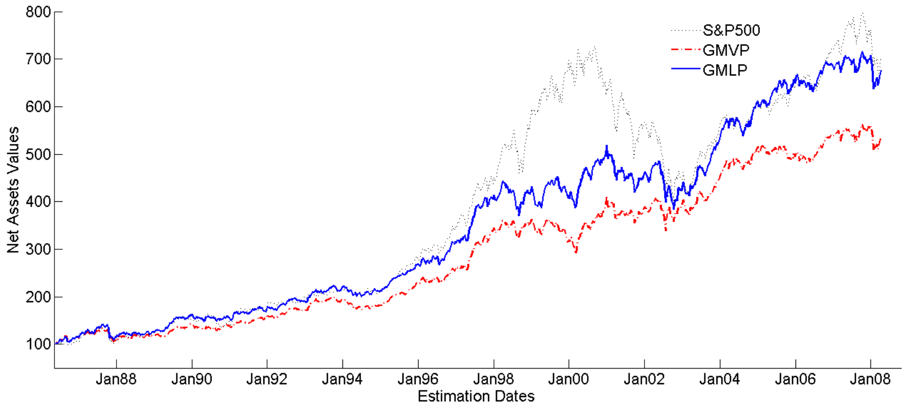

Figure 16: **Source** : *Reuters, net assets values in basis 100, out-of-sample GMVP and GMLP, 260 periods for the sample window, 1142 periods of estimation window, from 65 assets of the S&P500 index corresponding to the sectorial groups of deviating eigenvalues, no completion need, from 05/22/1981 to 04/11/2008, weekly frequency, computation by the authors.*

Figure 17: Volatility of GMVP and GMLP

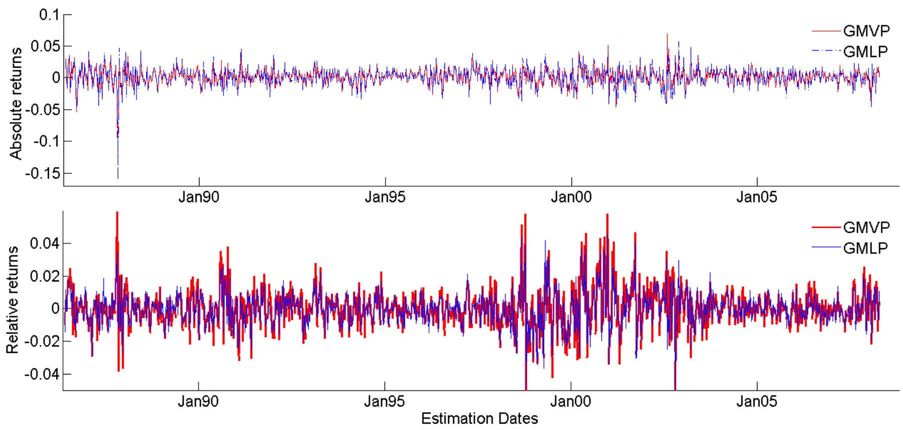

Figure 17: **Source** : *Reuters, volatility on the top and relative volatility bottom, out-of-sample GMVP and GMLP, 260 periods for the sample window, 1142 periods of estimation window, from 65 assets of the S&P500 index corresponding to the sectorial groups of deviating eigenvalues, no completion need, from 05/22/1981 to 04/11/2008, weekly frequency, computation by the authors.*

Figure 18: Effective Size and Turnover of GMVP and GMLP

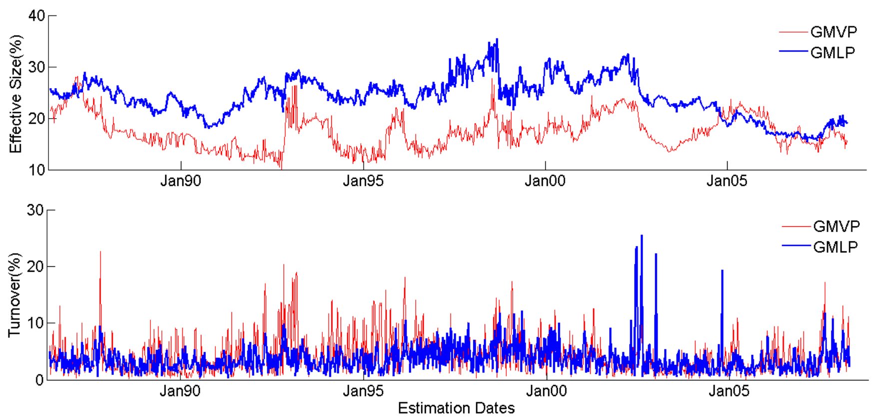

Figure 18: **Source** : *Reuters, effective size on the top and turnover bottom, of the out-of-sample GMVP and GMLP, 260 periods for the sample window, 1142 periods of estimation window, from 65 assets of the S&P500 index corresponding to the sectorial groups of deviating eigenvalues, no completion need, from 05/22/1981 to 04/11/2008, weekly frequency, computation by the authors.*

8.3 Formulas of the Statistics

Annualized Standard Deviation: ASD

$$ASD = \left[\frac{1}{T-1} \sum_{i=1}^{T} (x_i - \bar{x})^2 \right]^{1/2} * \sqrt{f}$$

where x_i denotes the portfolio returns, $\bar{x}$ denotes the sample mean returns, f the estimation's frequency, and T the size of the estimation period.

Annualized Mean Return: AMR

$$AMR = (1 + \bar{x})^f - 1$$

Sharpe Ratio: SR

$$SR = \frac{AMR}{ASD}$$

Tracking Error: TE

$$TE = \left[\frac{1}{T-1} \sum_{i=1}^{T} (y_i - \bar{y})^2 \right]^{1/2} * \sqrt{f}$$

where y_i denotes the difference between the portfolio returns and the market index, $\bar{y}$ denotes the corresponding sample mean returns.

Effective Size: ES

$$ES = \frac{1}{N \left(\sum_{j=1}^{N} \left(\mathbf{w}_{i,j}^* \right)^2 \right)}$$

where $\mathbf{w}_{i,j}^*$ denotes the optimal allocation for asset j at the date i, and N denotes the number of assets in the investment universe.

Turnover: TR

$$TR = \frac{1}{2} \sum_{j=1}^{N} \left| \mathbf{w}_{i+1,j}^* - \mathbf{w}_{i,j}^* \right|$$

where $\mathbf{w}_{i+1,j}^*$ denotes the optimal allocation at the date $i+1$ and $\mathbf{w}_{i,j}^*$ the optimal portfolio at the date i for asset j.